STATISTICAL INFERENCE

Reinhard Viertl
Institute of Statistics and Probability Theory, Vienna University of Technology, Austria.

Keywords: Bayesian statistics, classical statistics, confidence regions, data, data quality, estimation, fuzziness, likelihood function, nonparametric models, parametric models, semiparametric models, standard statistics, statistical decisions, stochastic models, sufficiency, testing

Contents

Summary: From descriptive statistics to statistical inference

Preliminary statistical data analysis is the first attempt to condense data and to describe relations between different variables. This is done without probability models, whereas the next step is to look for a stochastic model that describes the observed data in order to understand a phenomenon and to predict future values. These stochastic models are often determined by so-called *parameters*. The parameters are to be determined (estimated) from the observed data.

Another problem is to decide whether a certain stochastic model is appropriate for a real phenomenon. Decision rules for this are called *statistical tests.* Also of importance is the decision whether the underlying probability distribution of two data series can be considered to be identical. These and other related questions concerning decision support are the subject of statistical inference.

Statistical inference is the science of making justifiable conclusions concerning stochastic phenomena based on probability models and observations of the involved quantities.

1. Introduction

Complementary to descriptive statistical methods, the application of probability in the description of real phenomena is of increasing importance in the applied sciences, because there are no deterministic laws for many important processes. The up to date description of non-deterministic quantities is by *stochastic quantities X* and their probability distributions *P*. See the topic-level contribution *Probability* and the articles in this topic.

In statistical inference, probability models are used as the basis for the analysis and interpretation of data. In order to find the probability distribution of a stochastic quantity *X,* the observed data are considered to form a so-called *sample* of the stochastic quantity *X*. A sample is a finite sequence $X_1,\cdots,X_n$ of observations of *X* that have the same probability distribution as *X* and are statistically independent.

Statistical inference is the process of making mathematically sound conclusions concerning underlying probability models. In contrast to probability theory, which is a deductive mathematical theory, statistical inference is an inductive process in order to find suitable probability models to describe mathematically the laws behind phenomena that are observed with the help of data.

Besides modeling the real situation as correctly as possible, such models have unknown quantities that determine the probability models. These quantities are called *parameters*. In general a—possibly vector valued—parameter is denoted by θ and the set ob possible values for the parameter is denoted by Θ. Since Θ is a subset of the Euclidian space $\mathbb{R}^k$, Θ is called *parameter space*.

Example 1: Let the stochastic quantity *X* describe the number of point events in a time period of fixed length. Then *X* may be described by a Poisson distribution $Poi(\theta)$ which has a one-dimensional parameter $\theta > 0$. Compare the theme-level contribution *Probability and Statistics.* In this example the parameter space Θ is equal to the infinite interval $(0,\infty)$ of all positive real numbers. An inference problem is to estimate the true parameter value using observed numbers of point events in time intervals of the fixed considered length.

In looking for the underlying probability distribution of a stochastic quantity the concept of a *stochastic model* is helpful. A stochastic model is a family $\mathcal{P}$ of probability distributions that are suitable for the stochastic quantity *X* under consideration. Formally this is denoted by $X \sim P, \quad P \in \mathcal{P}$,

where $X \sim P$ means the probability distribution of *X* is *P*.

The next question is whether the probability distribution of *X* belongs to a subset $\mathcal{H}$ of the family $\mathcal{P}$ of all possible probability distributions. Such a subset $\mathcal{H}$ of $\mathcal{P}$ is called a *statistical hypothesis.*

Example 1 (continued): A related statistical hypotheses would be "the parameter of *X* is greater than 5." This would be denoted by the hypothesis

$\mathcal{H} : \theta \geq 5$,

and is the set of all Poisson-distributions whose parameter is not less than 5.

In general statistical hypotheses are assumptions about stochastic quantities. These can concern one or more stochastic quantities. In order to decide if a statistical hypothesis is acceptable, *statistical tests* were developed. The concept of a statistical test is explained in Section 0 of this article.

Another question in Example 1 is the following: For a given high probability $1-\alpha$, typically $1-\alpha = 0.95$ or 0.99, a subset $C_{1-\alpha}$ of the set Θ of all possible parameters is required, in which the true parameter is contained with probability $1-\alpha$. This set $C_{1-\alpha}$ is called the *confidence region* for the true parameter. The number $1-\alpha$ is called *coverage probability*.

All the procedures concerning the above mentioned problems belong to statistical inference.

In principle there are four categories of statistical inference procedures:

- point estimations for stochastic models and parameters
- confidence estimations for parameters of stochastic models
- statistical tests for statistical hypotheses concerning stochastic models and random phenomena, and
- optimal statistical decisions under uncertainty when loss or utility is involved.

An essential point of statistical inference is to learn from data at hand, not primarily to develop a formal theory.

Statistical inference procedures depend on the type of data which are given. The types of primary interest are:

(i) categorical data
(ii) ordinal data
(iii) interval data
(iv) ratio data.

These different types of data are of increasing specificity. *Categorical data* have no ordering scheme between their possible values. Examples are color, race, gender, and nationality. *Ordinal data* have a natural order in their possible values but no natural distance between them. Examples are rankings, quality classes, and examination marks. *Interval data* have the additional property that distances between data points can reasonably be defined. The origin is arbitrary. Examples are calendar data, time measurements, and temperature measurements. *Ratio data* are the most structured type. They are represented by numbers and there is a well-defined origin. Therefore it is reasonable to calculate ratios of two observations, because they do not depend on the measurement unit. Examples are length, age of a person, income, and mass.

For metric data (which are represented by numbers) there is a secondary classification into *discrete* and *continuous* data. For *discrete data* the set of possible values is at most countable (i.e. they can be identified with a sequence $a_1, a_2, \cdots$ of numbers) with no accumulation point. Usually the set of possible values is a subset of the set of all integers. *Continuous data* can assume all possible real numbers in an interval. This interval can be of finite length or infinite.

Another classification of data types is by dimensionality. Data are termed *univariate* if to every considered unit only one value—for metric data or interval data only one number—is taken. *Multivariate* data are those where different values are observed for every unit under consideration. For multivariate data also the dependence structure between the different observed data quantities is important.

Statistical inference can be considered as the route from data to stochastic models using decisions based on probability theory.

2. Parametric and Nonparametric Inference

Statistical inference can be classified in different ways. One classification is by the assumptions concerning the underlying probability models.

2.1. Parametric Inference

Where the family $\mathcal{P}$ of possible probability distributions for a stochastic quantity X is characterized by a finite number of parameters $\theta_1, \cdots, \theta_k$, these parameters together are called the *parameter vector* or *parameter* $\theta = (\theta_1, \cdots, \theta_k)$ and the family $\mathcal{P}$ is given by a so-called *parametric family*

$$\mathcal{P} = \{ P_\theta, \theta \in \Theta \}, \qquad (1)$$

where Θ denotes the set of all possible values for θ, and Θ is called the *parameter space.*

If the probability distribution of a stochastic quantity X can be assumed to be from a parametric family of probability distributions the corresponding stochastic model

$$X \sim P_\theta, \theta \in \Theta$$

is called a *parametric stochastic model.*

Inference procedures for such stochastic models can be reduced to inference procedures concerning the parameter θ of the model. In this case the corresponding statistical inference problems belong to so-called *parametric inference.*

Examples are all inference procedures for normal distributed stochastic quantities. In this case the parameter of the model is two-dimensional, i.e.

$$\theta = \left(\mu, \sigma^2 \right)$$

and the stochastic model is denoted by

$$X \sim N\left(\mu, \sigma^2 \right), \mu \in \mathbb{R}, \sigma^2 > 0,$$

where $N\left(\mu, \sigma^2 \right)$ denotes the normal distribution with expectation μ and variance σ^2. This distribution has the probability density (compare the theme level contribution *Probability and Statistics*).

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for all} \quad x \in \mathbb{R}. \tag{2}$$

Statistical inference is based on observations of the stochastic quantity and tries to find the best possible model to describe the stochastic quantity being considered. In case of a parametric model, inference procedures center on the parameter θ of the model. Details are given in Sections 0 and 0 of this article.

The use of parametric stochastic models can be justified if certain conditions are fulfilled. Examples are the Poisson process (compare the article *Construction of Random Functions and Path Properties*), and others are sums of a large number of independent stochastic quantities where the central limit theorem applies (compare the article *Limit Theorems of Probability Theory*).

2.2. Nonparametric Inference

Where we cannot assume a parametric model for a stochastic quantity, the corresponding inference methods belong to so-called *nonparametric inference*. Details of such inference procedures are given in Section 0 of this article.

2.3. Semiparametric Inference

Semiparametric inference procedures were developed in the late twentieth century. They are stochastic models that contain a finite dimensional parameter and unknown distribution functions that do not belong to a parametric family.

3. Sufficiency and Information

Sampling from a stochastic quantity X is carried out in order to obtain information about the distribution P of X. If a sample $X_1,\cdots,X_n$ of X is taken, an important question is: Can the sample be condensed without losing information about the distribution P of X ?

In general this is complicated, but for special situations an answer can be given. There are different definitions of sufficiency of a function $\mathsf{S}\left(X_1,\cdots,X_n\right)$ of the sample depending on the underlying paradigm, that is, classical statistics or Bayesian statistics. In many important practical cases the results coincide.

The essential point of sufficiency of a statistic $S=\mathsf{S}\left(X_1,\cdots,X_n\right)$ is that it is sufficient to know the value $s=\mathsf{S}\left(x_1,\cdots,x_n\right)$ for further statistical analysis. Such sufficient statistics can also be stochastic vectors.

Example: Let X denote the waiting time in a service system and have the exponential distribution Ex_θ with probability density

$$f\left(x\,|\,\theta\right)=\theta\cdot e^{-\theta x}I_{(0,\infty)}\left(x\right),\qquad \theta\in\Theta=\left(0,\infty\right) \tag{3}$$

where $I_A\left(\cdot\right)$ denotes the *indicator function* of the subset A of $\mathbb{R}$, i.e.

$$I_A\left(x\right)=\begin{cases}1 & \text{for} \quad x\in A\\ 0 & \text{for} \quad x\notin A.\end{cases} \tag{4}$$

In order to estimate the unknown parameter θ from the sample $X_1,\cdots,X_n$ of X it can be proved that the statistic

$$S=\mathsf{S}\left(X_1,\cdots,X_n\right)=\sum_{i=1}^{n}X_i \tag{5}$$

is sufficient. For details see Section 0 and Section 0.

3.1. The Likelihood Function

Let the stochastic quantity *X* with outcome space *M* have a discrete probability distribution with point probability $p(\cdot|\theta)$ depending on a parameter θ from the parameter space Θ. Having a sample $X_1,\cdots,X_n$ of *X*, *i.e.* *n* independent and identically like *X* distributed stochastic quantities, the joint distribution of the stochastic vector $(X_1,\cdots,X_n)$ is an *n*-dimensional discrete distribution on the sample space M^n, and the joint point probabilities are given by

$$q(x_1,\cdots,x_n|\theta)=p(x_1|\theta)\,p(x_2|\theta)\cdots p(x_n|\theta) \quad \text{for all} \quad (x_1,\cdots,x_n)\in M^n \tag{6}$$

After the observations $X_1=x_1,\cdots,X_n=x_n$ are obtained the quantities $x_1,\cdots,x_n$ are fixed numbers. Therefore the point probabilities in equation (6) are a function of the parameter θ. Therefore in statistical inference the following notation is used:

$$\ell(\theta;x_1,\cdots,x_n)=\prod_{i=1}^{n}p(x_i|\theta), \qquad \theta\in\Theta \tag{7}$$

and this function $\ell(\theta;x_1,\cdots,x_n)$ is called the *likelihood function.*

Where a stochastic quantity *X* has a continuous probability distribution with probability density $f(\cdot|\theta)$, $\theta\in\Theta$ the joint probability density of a sample $(X_1,\cdots,X_n)$ is given by

$$g(x_1,\cdots,x_n|\theta)=f(x_1|\theta)\,f(x_2|\theta)\cdots f(x_n|\theta)$$

for all $(x_1,\cdots,x_n)\in M^n$. (8)

This function—considered as a function of the variable θ—is analogously denoted by

$$\ell(\theta;x_1,\cdots,x_n)=\prod_{i=1}^{n}f(x_i|\theta), \quad \theta\in\Theta \tag{9}$$

and is called the *likelihood function* for the continuous stochastic model.

The likelihood function is one of the basic functions in statistical inference, both in estimation and testing of statistical hypotheses. Moreover it allows us to characterize the sufficiency of a statistic in case of parametric stochastic models.

The following theorem gives a sufficient condition for sufficiency of a statistic.

Theorem 1: Let the stochastic model $X \sim P_\theta, \theta \in \Theta$ be discrete or continuous. For sample $X_1, \cdots, X_n$ of X a statistic $S = \boldsymbol{S}(X_1, \cdots, X_n)$ is sufficient for θ if the likelihood function can be factorized in the following way:

$$\ell(\theta; x_1, \cdots, x_n) = g(s(x_1, \cdots, x_n), \theta)\, h(x_1, \cdots, x_n) \quad \text{for all } \theta \in \Theta \quad \text{and all } (x_1, \cdots, x_n) \in M_X^n \tag{10}$$

where the two functions $g(\cdot,\cdot)$ and $h(\cdot)$ are both nonnegative, $h(\cdot)$ is free from θ, and $g(s(x_1, \cdots, x_n), \theta)$ depends on $x_1, \cdots, x_n$ only through the observed value $s(x_1, \cdots, x_n)$ of the statistic *S*.

4. Classical Statistical Inference

The basic assumption in classical statistical inference is that there is a true but unknown probability distribution for an observable stochastic quantity or stochastic vector.

Statistical inference in this case has to draw conclusions about the true distributions. These conclusions can be estimates for unknown parameters, or tests of hypotheses concerning stochastic quantities and their probability distributions. In looking for estimates there are two different approaches: *point estimators* and *confidence regions.*

-
-
-

Bibliography

Berger J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Verlag. [Comprehensive text on mathematical aspects of Bayesian inference and decision analysis.]

Bromek T. and Pleszczynska E. (eds.) (1991). *Statistical Inference. Theory and Practice*. Dordrecht: Kluwer Academic. [Contains theoretical and applied chapters on statistical inference written by different authors, as well as applications from life sciences, and many references.]

Gibbons J.S. and Chakraborti S. (1992). *Nonparametric Statistical Inference*. New York: Marcel Dekker. [Comprehensive text on nonparametric methods of statistics.]

Kalbfleisch J.G. (1985). *Probability and Statistical Inference, Vol. 2: Statistical Inference*, New York: Springer Verlag. [A well written introduction to the principles of statistical inference]

Mardia K.V., Kent J., and Bibby, J.M. (1979). *Multivariate Analysis*. London: Academic Press. [High-level text on the foundations of statistical inference for vector-valued data.]

Mukhopadhyay N. (2000). *Probability and Statistical Inference*. New York: Marcel Dekker. [Standard text on mathematical aspects of probability and statistical inference.]

Viertl R. (1996). *Statistical Methods for Non-Precise Data*. Boca Raton, Florida: CRC. [Basic text on the description and statistical analysis of non-precise data.] 191 pp.

Biographical Sketch

Reinhard Viertl was born in 1946, at Hall in Tyrol, Austria. He studied civil engineering and engineering mathematics at the Technische Hochschule in Vienna, and received his Dipl.-Ing. degree in engineering mathematics in 1972 and Doctorate of Engineering Science degree in 1974. He was appointed assistant at the Vienna Technische Hochschule and promoted to University Docent in 1979. He worked as a research fellow and visiting lecturer at the University of California, Berkeley, from 1980 to 1981, and visiting Docent at the University of Klagenfurt, Austria in the winter of 1981–1982. Since 1982 he has been full Professor of Applied Statistics at the Department of Statistics, Vienna University of Technology, and was also visiting professor at the Department of Statistics, University of Innsbruck, Austria from 1991 to 1993. He has served as organizer of a number of scientific conferences

He is a fellow of the Royal Statistical Society, London, held the Max Kade fellowship in 1980, is a founder of the Austrian Bayes Society, and member of the International Statistical Institute. He served as president of the Austrian Statistical Society from 1987 to 1995, and was invited to become a member of the New York Academy of Sciences in 1998.

His publications include: *Statistical Methods in Accelerated Life Testing* (1988), *Introduction to Stochastics* (in German, 1990), *Statistical Methods for Non-Precise Data* (1996), and over 90 scientific papers in algebra, probability theory, accelerated life testing, regional statistics, and statistics with non-precise data. He edited *Probability and Bayesian Statistics* (1987) and *Contributions to Environmental Statistics* (in German, 1992). He was co-editor of *Mathematical and Statistical Methods in Artificial Intelligence* (1995), and two special volumes of journals. He is editor of the publication series of the Vienna University of Technology and a member of the editorial board of a number of scientific journals.